UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/672,248	09/29/2003	Monika Henzinger	0026-0043	8702

44989 7590 03/08/2010

HARRITY & HARRITY, LLP
11350 Random Hills Road
SUITE 600
FAIRFAX, VA 22030

EXAMINER	
RUTLEDGE, AMELIA L	
ART UNIT	**PAPER NUMBER**
2176	

MAIL DATE	DELIVERY MODE
03/08/2010	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS
AND INTERFERENCES

Ex parte MONIKA HENZINGER and MATTHIAS RUHL

Appeal 2009-003375
Application 10/672,248
Technology Center 2100

Decided: March 8, 2010

Before JAMES D. THOMAS, LEE E. BARRETT, and
ST. JOHN COURTENAY III, *Administrative Patent Judges*.

COURTENAY, *Administrative Patent Judge*.

DECISION ON APPEAL

Appellants appeal under 35 U.S.C. § 134(a) from the Examiner's rejection of claims 1-8, 10, 12-21, 23-26, and 28-30. Claims 9, 11, 22, 27, and 31 have been cancelled. We have jurisdiction under 35 U.S.C. § 6(b).

We reverse.

STATEMENT OF THE CASE

INVENTION

Appellants' invention relates generally to content retrieval on the World Wide Web. More particularly, the invention on appeal is directed to automated web crawling. (Spec., para. [0002]).

ILLUSTRATIVE CLAIMS

1. A method for crawling documents comprising:

receiving a uniform resource locator (URL);

receiving at least two different copies of a document associated with the URL; and

determining whether a web site corresponding to the URL uses session identifiers based on a comparison of URLs that are within the document and that change between the at least two different copies of the document, where the web site is determined to use session identifiers when a portion of the URLs that change between the at least two different copies of the document is greater than a threshold.

10. A method for identifying web sites that use session identifiers comprising:

downloading at least two different copies of at least one document from a web site;

extracting uniform resource locators (URLs) from the two different copies of the web document;

comparing the extracted URLs of the two different copies of the document; and

determining whether the web site uses session identifiers when the comparison indicates that at least a portion of the URLs change between the two different copies.

15. A device comprising:

a spider component configured to crawl web documents associated with at least one web site; and

a session identifier component configured to determine whether the web site uses session identifiers based on a comparison of a portion of uniform resource locators (URLs) that change between different copies of at least one web document downloaded from the web site.

PRIOR ART

Galai	WO 03/017023 A2	Feb. 27, 2003
DaCosta	US 6,665,658 B1	Dec. 16, 2003

THE REJECTION

The Examiner rejected claims 1-8, 10, 12-21, 23-26, and 28-30 under 35 U.S.C. § 103(a) as unpatentable over the combination of Galai and DaCosta.

CONTENTIONS BY APPELLANTS

Appellants contend, *inter alia*, that "[c]omparing a web page for similarity in content or visual similarity, as disclosed by Galai, . . . cannot be said to disclose or suggest determining whether a web site corresponding to

a URL uses session identifiers 'based on a comparison of URLs that are within the document and that change between the at least two different copies of the document . . . .'" (App. Br. 8, ¶3).

THE EXAMINER'S RESPONSE

The Examiner disagrees. The Examiner proffers that "it would have been obvious to one of ordinary skill in the art at the time of the invention that the same [Galai's] threshold could be used to determine the difference between URLs for web pages, since the difference threshold [as claimed] *would have been the inverse or opposite* of [Galai's] *similarity threshold number*." (Ans. 5, ¶2, last four lines, emphasis added).

ISSUE

Based upon our review of the administrative record, we have determined that the following issue is dispositive in this appeal:

> Under § 103, would the combination of Galai and DaCosta have taught or suggested determining whether a web site corresponding to a URL uses session identifiers based on a comparison of URLs that are within the document and that change between the at least two different copies of the document?

PRINCIPLES OF LAW

"What matters is the objective reach of the claim. If the claim extends to what is obvious, it is invalid under § 103." *KSR Int'l Co. v. Teleflex, Inc.*, 550 U.S. 398, 419 (2007). To be nonobvious, an improvement must be

"more than the predictable use of prior art elements according to their established functions." *Id.* at 417.

FINDINGS OF FACT

1. Galai teaches that "the comparison function of the present invention preferably checks for similarity in content and more preferably produces a similarity level, which is the likelihood of the two Web pages to have the same content. If this value exceeds a certain threshold, then most preferably the removed parameter is considered to be redundant." (P. 27, ll. 18-22).
2. Galai teaches that "[a]ccording to preferred embodiments of the present invention, the level of similarity is determined according to visual similarity. Visual similarity is preferably determined according to two different types of parameters. A first type of parameter is based upon content of the document, such as text and/or images for example. A second type of parameter is based upon visual layout characteristics of the document, such as the presence of one or more GUI (graphical user interface) gadgets or the location of text and/or images, for example." (P. 27, l. 23 through p. 28, l. 6).
3. DaCosta teaches determining if a website listed on a manually created URL site list is an interactive web site that sets "cookies" by performing a "GET method" (as defined by the hypertext transfer protocol, HTTP) to download the content of the website. (Col. 6, ll. 22-30).

ANALYSIS

Independent claims 1, 10, 15, 21, and 26

We decide the question of whether the combination of Galai and DaCosta would have taught or suggested determining whether a web site corresponding to a URL uses session identifiers based on a comparison of URLs that are within the document and that change between the at least two different copies of the document. (*See* commensurate limitations recited in each of independent claims 1, 10, 15, 21, and 26).

After considering the evidence before us, and the respective arguments on both sides, we conclude that the Examiner's proffered combination of Galai and DaCosta would not have fairly rendered obvious Appellants' claimed invention. In particular, we agree with Appellants' principal argument that Galai's comparison function that checks for similarity in content (as described by Galai at page 21, lines 4-7), does not teach nor fairly suggest determining whether a web site corresponding to a URL uses session identifiers based on a comparison of URLs that are within the document and that change between the at least two different copies of the document, within the meaning of Appellants' independent claims on appeal. (App. Br. 8-9).

Specifically, we note that Galai teaches "the comparison function of the present invention preferably checks for similarity in content and more preferably produces a similarity level, *which is the likelihood of the two Web pages to have the same content.*" (FF 1, emphasis in original). In contrast,

Appellants' claimed invention looks for changes or differences in URLs between two different copies of a web page document and thus compares URLs embedded within two documents.

We find unpersuasive the Examiner's gap-filling reasoning that "it would have been obvious to one of ordinary skill in the art at the time of the invention that the same [Galai's] threshold could be used to determine the difference between URLs for web pages, since the difference threshold [as claimed by Appellants] *would have been the inverse or opposite* of [Galai's] *similarity threshold number*." (Ans. 5, ¶2, last four lines, emphasis added). We observe that Galai's similarity level is clearly described as "the likelihood of the two Web pages to have the same content." (FF 1, underline added). As discussed *supra*, Appellants' claimed invention looks for changes or differences in URLs between two different copies of a web page document. If anything, the Examiner has laid the foundation for a "teaching away" argument.

Moreover, we find Galai teaches that the comparison to determine the level of similarity between two web pages is performed according to the level of visual similarity of the web pages, based on two specific types of parameters. (FF 2). Galai describes the first type of parameter as being based upon the content of the document, such as text and/or images. (*Id.*). Galai describes second type of parameter is based upon the visual layout characteristics of the document, such as the presence of one or more GUI (graphical user interface) gadgets, or the location of text and/or images. (*Id.*). Thus, we find Galai compares visual elements (instead of the URLs

compared by Appellants' claimed invention), and also looks for similarities between the visual elements instead of looking for changes in URLs (i.e., differences), as claimed by Appellants.

For at least the aforementioned reasons, we find Galai's comparison that looks for similarities between visual or layout elements of two web pages does not teach nor fairly suggest determining whether a web site corresponding to a URL uses session identifiers based on a *comparison of URLs that are within the document and that change between the at least two different copies of the document.* (*See* commensurate limitations recited in each of independent claims 1, 10, 15, 21, and 26).

While the Examiner looks to the secondary DaCosta reference as purportedly teaching that URLs are compared for the specific purpose of determining whether the web site uses session identifiers (Ans. 5-6), we find the portion of DaCosta relied on by the Examiner performs no comparison of URLs between different copies of at least one web page document, within the meaning of Appellants' independent claims. Instead, DaCosta determines if a website listed on a manually created URL site list is an interactive web site that sets "cookies" by performing a "GET method" (as defined by the hypertext transfer protocol (HTTP)) to download the content of the website. Therefore, we find DaCosta fails to overcome the deficiencies of Galai.

Accordingly, we reverse the Examiner's obviousness rejection of independent claims 1, 10, 15, 21, and 26. Because we have reversed the Examiner's rejection of each independent claim on appeal, we also reverse the Examiner's obviousness rejection of each dependent claim on appeal.

CONCLUSION

The Examiner's proffered combination of Galai and DaCosta does not teach nor fairly suggest determining whether a web site corresponding to a URL uses session identifiers based on a comparison of URLs that are within the document and that change between the at least two different copies of the document. (*See* commensurate limitations recited in each of independent claims 1, 10, 15, 21, and 26).

ORDER

We reverse the Examiner's decision rejecting claims 1-8, 10, 12-21, 23-26, and 28-30 under 35 U.S.C. § 103(a).

REVERSED

pgc

HARRITY & HARRITY, LLP
11350 Random Hills Road
SUITE 600
FAIRFAX, VA 22030